

Structural Analysis Quick Start

An NCBI Mini-Course

A protein domain is considered to be a distinct functional and/or structural unit. A domain in a structural context refers to a segment of a polypeptide chain that can fold into an independent three dimensional structure. It may interact with other domains of the protein or may simply be joined to other domains by a polypeptide chain. A domain in a sequence context refers to a long sequence pattern that is shared by other proteins having a common evolutionary origin. A domain may include all of the protein sequence or a part of it. A conserved domain is a recurring unit in molecular evolution whose extents can be determined by sequence and structure analysis.

The Conserved Domain Database (CDD) contains domains derived from the Smart, Pfam and Clusters of Orthologous Groups (COGs) databases. Conserved domains can be represented as multiple sequence alignments. Source alignments are processed by NCBI as follows:

- Sequences in the alignment for which a link can not be provided to a protein in Entrez are removed.
- If possible, a closely related sequence with a known structure is substituted.
- A representative sequence, preferably with a structure link, is chosen from among those in the alignment.
- A consensus sequence is made.
- A position-specific scoring matrix (PSSM) is constructed.

The Conserved Domain search (CD-search) compares a protein sequence to the PSSMs in the CDD database to identify conserved domains within it and to identify a 3-D modeling template. Since the PSSMs are the "subject", instead of the query as in PSI-Blast, the CD-search is a form of Reverse Position-Specific Blast (RPS-Blast).

The Conserved Domain Architecture Retrieval Tool (CDART) can be used to identify proteins containing the domain(s) present in the query sequence. Conserved domain(s) present in all sequences within Entrez proteins are identified using CD-search during routine NCBI processing. These pre-computed results are accessed through CDART.

The Vector Alignment Search Tool (VAST) is a computer algorithm developed at NCBI to detect similar protein 3-dimensional structures. The "structure neighbors" for every structure in NCBI's Molecular Modeling DataBase (MMDB)

are pre-computed. These neighbors can be used to identify distant homologs that cannot be recognized by sequence comparison alone. A VAST-search can be used for determining the structure neighbors for recently solved structures not yet in MMDB.

Cn3D is a helper application for web browsers to view 3-dimensional structures from NCBI's Entrez retrieval service. Cn3D runs on Windows, Macintosh, and Unix. Cn3D simultaneously displays structure, sequence, and alignment, and now has powerful annotation and alignment editing features.

In this course, we will learn to

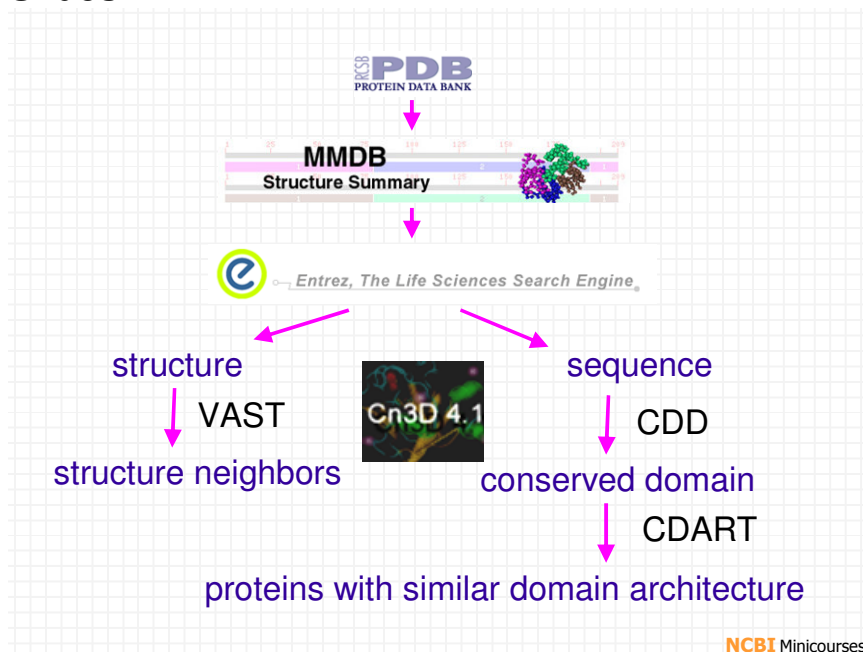
- Identify a conserved domain present in the query protein using **CDD**
- Search for other proteins containing similar domain(s) using **CDART**
- Explore a 3D modeling template for the query sequence using **CDD**
- Find similar structures using **VAST**
- Visualize and annotate the 3D protein structures using **Cn3D**

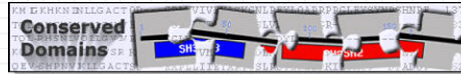
The remainder of the handout includes the introductory slides and the screen shots of the exercise demonstrated in Problem 1.

URL: <http://www.ncbi.nlm.nih.gov/Class/minicourses/quickstructure.html>

Course developed by: Dr. Medha Bhagwat (bhagwat@ncbi.nlm.nih.gov)

Slides





Conserved Domains

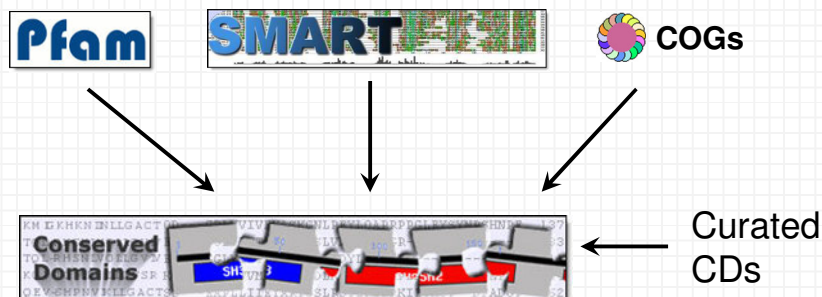
- recurring unit in molecular evolution;
extent can be determined by
sequence and structure analysis;
- performs a defined function;
- represented as a multiple local sequence
alignment of proteins containing the domain
(position specific scoring matrix, pssm).

CDD Search page:

<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>

NCBI Minicourses

Conserved Domain Database

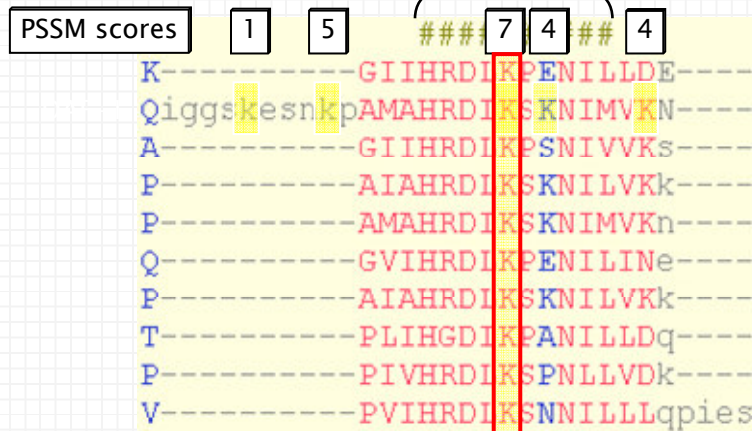


- A database of position-specific scoring matrices (PSSM)
- CD-Search can be used to search against the PSSMs
- Manual curation of CDs is ongoing (cd12345.version)

NCBI Minicourses

Position-Specific Score Matrix

Serine/Threonine protein kinases
catalytic loop



NCBI Minicourses

Position-Specific Score Matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
435 K	-1	0	0	-1	-2	3	0	3	0	-2	-2	1	-1	-1	-1	-1	-1	-1	-1	-2
436 E	0	1	0	2	-1	0	2	-1	0	-1	-1	0	0	0	-1	0	0	-1	-1	-1
437 S	0	0	-1	0	1	1	0	1	1	0	-1	0	0	0	2	0	-1	-1	0	-1
438 N	-1	0	-1	-1	1	0	-1	3	3	-1	-1	1	-1	0	0	-1	-1	1	1	-1
439 K	-2	1	1	-1	-2	0	-1	-2	-2	-1	-2	5	1	-2	-2	-1	-1	-2	-2	-1
440 P	-2	-2	-2	-3	-2	-2	-2	-2	-1	-2	-1	0	-3	7	-1	-2	-3	-1	-1	-1
441 A	3	-2	1	-2	0	-1	0	1	-2	-2	-2	0	-1	-2	3	1	0	-3	-3	0
442 M	-3	-4	-4	-4	-3	-4	-4	-5	-4	7	0	-4	1	0	-4	-4	-2	-4	-1	2
443 A	4	-4	-4	-4	0	-4	-4	-3	-4	4	-1	-4	-2	-3	-4	-1	-2	-4	-3	4
444 H	-4	-2	-1	-3	-5	-2	-2	-4	10	-6	-5	-3	-4	-3	-2	-3	-4	-5	0	-5
445 R	-4	8	-3	-4	0	-1	-2	-3	-2	-5	-4	0	-3	-2	-4	-3	-3	0	-4	-5
446 D	-4	-4	-1	8	-6	-2	0	-3	-3	-5	-6	-3	-5	-6	-4	-2	-3	-7	-5	-5
447 I	-4	-5	-6	-6	-3	-4	-5	-6	-5	3	5	-5	1	1	-5	-5	-3	-4	-3	1
448 K	0	0	1	-3	-5	-1	-1	-3	-3	-5	-5	7	-4	-5	-3	-1	-2	-5	-4	-4
449 S	0	-3	-2	-3	0	-2	-2	-3	-3	-4	-4	-2	-4	-5	2	6	2	-5	-4	-4
450 K	0	3	0	1	-5	0	0	-4	-1	-4	-3	4	-3	-2	2	1	-1	-5	-4	-4
451 N	4	-3	8	-1	-5	-2	-2	-3	-1	-6	-6	-2	-4	-5	-4	-1	-2	-6	-4	-5
452 I	3	-5	-5	-6	0	-5	-5	-6	-5	6	2	-5	2	-2	-5	-4	-3	-5	-3	3
453 M	-4	-4	-6	-6	-3	-4	-5	-6	-5	0	6	-5	1	0	-5	-4	-3	-4	-3	0
454 V	-3	-3	-5	-6	-3	-4	-5	-6	-5	3	3	-4	2	-2	-5	-4	-3	-5	-3	5
455 K	-2	1	1	4	-5	0	-1	-2	1	-4	-2	4	-3	-2	-3	0	-1	-5	-2	-3
456 N	1	1	3	0	-4	-1	1	0	-3	-4	-4	3	-2	-5	-2	2	-2	-5	-4	-4
457 D	-3	-2	5	5	-1	-1	1	0	-5	-4	0	-2	-5	-1	0	-2	-6	-4	-5	-5
458 L	-3	-1	0	-3	0	-3	-2	3	-4	-2	3	0	1	1	-2	-2	-3	5	-1	-3

catalytic
loop

NCBI Minicourses

Problem 1

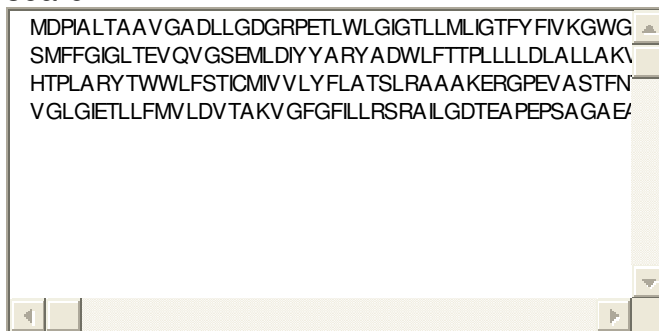
In this problem, we will follow these steps:

- A. Identify conserved domain(s) present in a protein.
- B. Search for other proteins containing similar domain(s).
- C. Explore a 3D modeling template for the query sequence.
- D. Find distant sequence homologs that may not be identified by BLAST.

NCBI's Conserved Domain Search allows you to match your protein sequence to a library of conserved protein domains, generate a multiple sequence alignment based on this match, and explore 3D modeling templates for your sequence. Click on the CDD link provided below,

CDD

Paste the following protein sequence in the CD-Search query box and run the search.



MDPIALTA AVGADLLGDGRPETLWLGIGTLLMLIGTFYFIVKGWG
SMFFGIGLTEVQV GSEMLDIYARYADWLFTPLLLDLALLAKV
HTPLARYTWWLFSTICMIVVLYFLATSLRAAAKERGPVESTFN
VGLGIETLLFMVLDVTAKVGFGFILLRSRAILGDTEAPEPSAGAE/

- A. What is the domain present in this protein?
Obtain more information about the domain by searching in [NCBI's Bookshelf](#)
- B. Go back to the CD-Search results page. Obtain a list of proteins with similar domain architecture by clicking on the "Search for similar domain architectures" button. To display the records, click on the link to the sequences and from there on the "Look up Sequences in Entrez". Change the display from "Summary" to "FASTA".
- C. Go back to the CD-Search results page. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the graphic of the domain. Use the "Row Display" list box pull down menu to specify "up to 5" sequences and reformat sequence alignment. Extend the "Structure" display and invoke Cn3D with a display of a 3D modeling template and a multiple sequence alignment including your query sequence by pressing the "Show Structure" button.

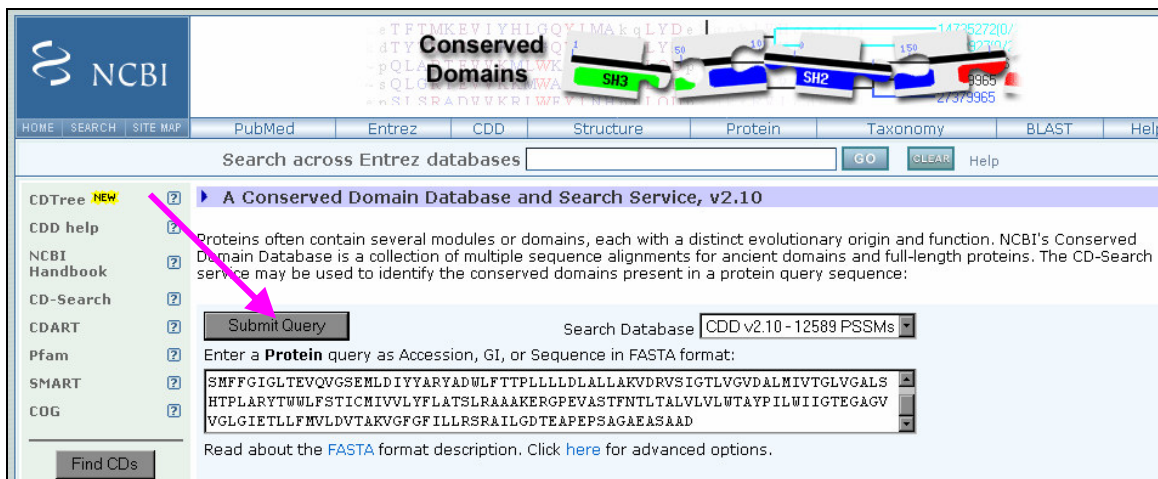
The structure of the *Halobacterium salinarum* halorhodopsin protein and its

sequence alignment with our query protein are displayed. For a better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a bacterial rhodopsin signature (FMVLDTAKVGF) where K is the retinal binding site. Identify these residues in the query protein and highlight the corresponding lysine residue in the halorhodopsin protein sequence.

Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click the Protein Side chains on. Click on the "Done" button. To zoom in, press z on the keyboard. Identify the cofactor near the lysine residue.

D. To obtain the structural neighbors for the halorhodopsin protein, first click on the structure entry link, 1E12_A, on the CD-Browser page. Then click Links → Structure on the top right, then on 1E12 again in the Entrez Structure page, and finally on the chain A graphic. To view neighbors with 1E12_A, select one or more of the check boxes next to the structure neighbors and view by clicking on the "View 3D Structure" button.

Screenshots



The screenshot displays the NCBI Conserved Domain Database (CDD) search interface. At the top, there is a protein structure visualization with domains SH3 and SH2 highlighted. Below this, a search bar is present with a 'GO' button. The main section is titled 'A Conserved Domain Database and Search Service, v2.10'. It includes a 'Submit Query' button and a search database dropdown set to 'CDD v2.10 - 12589 PSSMs'. The search results show a protein sequence: SMFFGIGLTEVQVGSEMLDIYYARYADWLF... and a description: Read about the FASTA format description. Click here for advanced options.

NCBI

Conserved Domains

Query sequence: [(local sequence)|c|Undefined_sequence]

☒ Concise Result ☐ Full Result ☐ Show Search Information ☒

Bac_rhodopsin

Descriptions

Title	Pssmid	Multi-Dom	E-value
Hpfpam01036, Bac_rhodopsin, Bacteriorhodopsin...	41106	No	1e-47

[Search for similar domain architectures](#)

CD Search Reference:

Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", *Nucleic Acids Res.* 32(W):327-331.

[Help](#) | [Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

NCBI

National Center for Biotechnology Information

National Library of Medicine National Institutes of Health

PubMed All Databases BLAST OMIM Books TaxBrowser Structure

Search All Databases for bacteriorhodopsin Go

What can NCBI do?

Hot Spots

- Assembly Archive
- Clusters of orthologous groups
- Coffee Break, Genes & Disease, NCBI Handbook
- Electronic PCR

NCBI

Bookshelf

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Books for bacteriorhodopsin Go Clear Save Search

Limits Preview/Index History Clipboard Details

About Entrez

Books

Overview

Using the books

Information for authors and publishers

Contact us

Mailing list

Display Books Show 20 Send to

All: 29 Figures: 11

11 items in Molecular Biology of the Cell. 4th ed.
 Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter.
 New York: Garland Publishing; c2002.

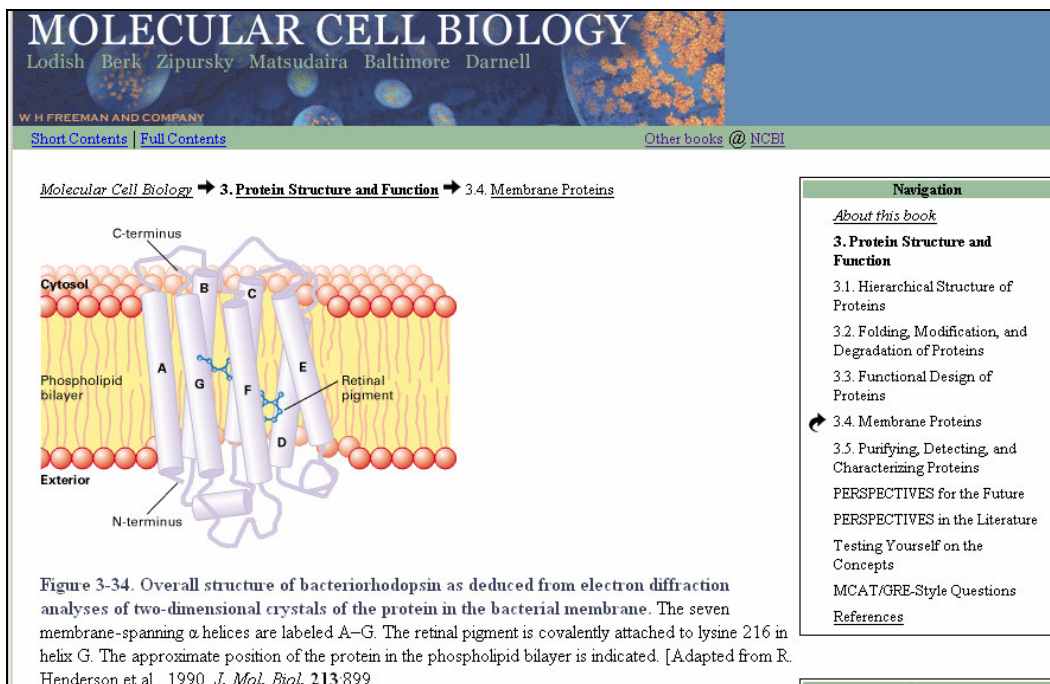
8 items in Biochemistry.
 Berg, Jeremy M.; Tymoczko, John L.; and Stryer, Lubert.
 New York: W. H. Freeman and Co.; 2002.

6 items in Molecular Cell Biology. 4th ed.
 Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E.
 New York: W. H. Freeman & Co.; c2000.

Many Integral Proteins Contain Multiple Transmembrane α Helices

Although [Figure 3-33](#) depicts glycophorin as a monomer with a single α helix spanning the bilayer, this protein is present in erythrocyte membranes as a dimer of two identical polypeptide chains. The two membrane-spanning α helices of glycophorin are thought to form a coiled-coil structure (see [Figure 3-9a](#)) stabilized by specific interactions between the amino acid side chains at the interface of the two helices. It is now known that many other transmembrane proteins contain two or more membrane-spanning α helices. For instance, the *bacterial photosynthetic reaction center (PRC)* comprises four subunits and several prosthetic groups, including four chlorophyll molecules. In this complex protein, three of the four subunits span the membrane; two of these subunits (L and M) each contain five membrane-spanning α helices (see [Figure 16-40](#)).

A large and important family of integral proteins is defined by the presence of seven membrane-spanning α helices. More than 150 such “seven-spanning” membrane proteins have been identified. This class of integral proteins is typified by *bacteriorhodopsin*, a protein found in a photosynthetic bacterium ([Figure 3-34](#)). Absorption of light by the retinal group attached to *bacteriorhodopsin* causes a conformational change in the protein that results in pumping of protons from the cytosol across the bacterial membrane to the extracellular space. The proton concentration gradient thus generated across the membrane is used to synthesize ATP, as discussed in [Chapter 16](#). Both the overall arrangement of the seven α helices in *bacteriorhodopsin* and the identity of most of the amino acids can be resolved by computer analysis of micrographs of two-dimensional crystals of the membrane-embedded protein taken at various angles to the electron beam.



NCBI

Conserved Domains

Query sequence: [(local sequence)|c|Undefined_sequence]

☒ Concise Result ☐ Full Result ☐ Show Search Information ☐

Bac_rhodopsin

Descriptions

Title	PssmId	Multi-Dom	E-value
[+]pfam01036, Bac_rhodopsin, Bacteriorhodopsin...	41106	No	1e-47

[Search for similar domain architectures](#)

NCBI

CDART: Conserved Domain Architecture Retrieval Tool

[New Query](#) [Overview](#) [PubMed](#) [Nucleotide](#) [Protein](#) [Structure](#) [Taxonomy](#)

[About CDART](#)

Query: **Bac_rhodop**

[Similar domain architectures](#)

559 Sequences
cellular organism
hypothetical prote

NCBI

CDART: Conserved Domain Architecture Retrieval Tool

[New Query](#) [Overview](#) [PubMed](#) [Nucleotide](#) [Protein](#) [Structure](#) [Taxonomy](#)

[About CDART](#)

Query: **Bac_rhodop**

[Similar domain architectures](#)

ZJHFA Halobacterium sal unnamed protein pr	none>
YP_001272404 Brevibacterium lact opsin, putative	none>
YP_001262778 Mecanopsis fische opsin, putative	none>
ZP_01616930 marine gamma prote bacteriorhodopsin	none>
YP_001240940 Coccidioides imit hypothetical prote	none>
Z121A Halobacterium sal Opsin R, Bacterior	none>
Z11XA Halobacterium sal Opsin R, Bacterior	none>
ABJ98286 uncultured bacteri putative protoorho	none>
ABJ98284 uncultured bacteri putative protoorho	none>

[Look Up Sequences in Entrez](#)

NCBI Entrez Protein

Search: Protein for [] Go Clear

Limits Preview/Index History Clipboard Details

Display: Summary Show: 20 Sort by: Relevance Send to: []

All: 557 Related Structures: 557

FASTA

1: GenPept
2: GI List
3: Graphics
4: TinySeq XML
5: INSDSeq XML
6: LinkOut
7: Related Sequences
8: Conserved Domain Links
9: 3D Domain Links
10: Gene Links
11: Genome Links
12: Genome Project Links
13: HomoloGene Links
14: Nucleotide Links
15: NIH cDNA clone links

bacteriorhodopsin [marine gamma proteobacterium HTCC2143]
gi|119476620|ref|ZP_01616930.1|[119476620]

NCBI Entrez Protein

Search: Protein for [] Go Clear

Limits Preview/Index History Clipboard Details

Display: FASTA Show: 20 Send to: []

Item 1 - 20 of 559

1: 2JAF Reports unnamed protein p...[gi|122920878]
>gi|122920878|pdb|2JAF|A unnamed protein product [Halobacterium salinarum]
MSITSPGVVDAGVLGAQSAARRENALLSSSLVWVVALAGIALLVFFVYMGRTIRPGRPRLIWGATLMIP
LVSISYLGSLTGVMIENPAGHALAGENVRSQWGRYLTWALSTPMLLALGLLADVDLGLFTVIAA
DGMCVTGLAAAMTTALLFRWAFYAISCAFFVVLVSALVTDWAASASSAGTAEIFDTLVLVVLVWLGY
PIVWAVGVEGLALVQSVGATSWAYSVDVFAKYVFAFILLRVVANNERTVAVAGQTLGTMSDD

2: XP_001272404 Reports opsin, putative [...[gi|121709381]
>gi|121709381|ref|XP_001272404.1| opsin, putative [Aspergillus clavatus NRRL 1]
MSHTKQPLPWPFTTSSVGPVPTVPGNEPIYQEIQVTKRALVWVTLMGVSSLVFYTLARLPKRVF
HTLVSIMTTISFITYLALATGSGMTWKHDSITRTHKHVPDTPQYFRQVWMLRYLNNFLTEPLGLINLSL
LSGLPGAHLVAIVADYIMLGSGLLGTGAGTSRRVWFTTISALGYLTYYHVGINGGAANNKDAQTRR
FFASLSGVTLVKALYPIALAAGPLAKMMNVNTETVIFAVDFITQGLGYWLIIAHDSPPGITLYADGF
WSGGIGNEGAIRINEEEDGA

3: XP_001262778 Reports opsin, putative [...[gi|119488656]
>gi|119488656|ref|XP_001262778.1| opsin, putative [Neosartorya fischeri NRRL 181]
MANPLRVVTLMLGSLVYFVTLARVPLSKRVFHTLVSINTVSFIVLALATGSGMAKHDLSLKHKH
VPDTPQYFRQVWMLRYLNNFLTEPLSLINLALVSLPGAHLVAIAADYVMLGSLGLTGFVGHSTRRV
WFTVSLGYLTYYHVAINGGAANNKDAQTRRFFASLSAVTLIVKVLVPIALAAGCLALRMNVDTETV
FAIYDITQGLGYWLIIAHDSAQGISLVYVDGFWSNGIGNEGAIRISEEDGA

4: ZP_01616930 Reports bacteriorhodopsin [...[gi|119476620]
>gi|119476620|ref|ZP_01616930.1| bacteriorhodopsin [marine gamma proteobacterium HTCC2143]
MTNLASDVPVGMFSLISNAAVATVFLIERDVSQWKTSLTVAGLVTLIAVHYFYNRDVWVATGE
TPTVRYIDULLTVPLLIEFYLLSAITKVPVGVFWLLAGSLIMLGAGFVGEVNPVYVSGFVVGMLG
WVWIMYIFLGEASKINAAAGNAIQKAYGAMRLLVTVGWAIYPIGYVGYTGSTDSATNLNVWVAVDL
WNKVAFLVIAAAVADSE

NCBI Conserved Domains

HOME SEARCH SITE MAP NewSearch PubMed Nucleotide Protein Structure CDD Taxonomy Help

Query sequence: [(local sequence)cl|Undefined_sequence]

Concise Result Full Result Show Search Information []

1 50 100 150 200 250

Bac_rhodopsin

Descriptions

Title	PssmId	Multi-Dom	E-value
[+]pfam01036, Bac_rhodopsin, Bacteriorhodopsin	PF00066	No	1e-47

Search for similar domain architectures

NCBI

Conserved Domains

pfam01036.12 Bac_rhodopsin, with user query added

[+] Links: Bacteriorhodopsin.

[+] Statistics:

[+] Structure:

Other Related Conserved Domains: C005524

Reformat Sequence Alignment Format: Compact Hypertext Row Display: Up to 5 Color Bits: 2.0 bits Type Selection: the most similar member

```

1E12_A      8 . [16] .LVFVYH. [1] .RTIRPGRPRLIWGATLMIPLVSISSYLGLLSGLTVGHIEMP. [11] .SQWGRYLTWALSTPMI 98
query      21 . [16] .FYFIVK. [1] .WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG. [5] .IYYARYADWLFITPLL 105
1UAZ_A     15 . [16] .FYFIVK. [1] .WGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG. [5] .IYYARYADWLFITPLL 99
1MOK_A     22 . [16] .LVFLVK. [1] .WGVSDPAKKFYAITTLVPAIAFTNLSMLLGYGLTMVFPF. [5] .IYYARYADWLFITPLL 106
gi 114809   34 . [16] .LVFVYH. [1] .RGLDDPRAKLIIVSTILVFPVSIASYTGLASGLTISVLEMP. [21] .TMWGRYLTWALSTPMI 134
gi 461609   29 . [16] .LVFVYH. [1] .RVNEDPRAQLIFVATLMVPLVSISSYTGLVSGLTISVLEMP. [11] .TPWGRYLTWALSTPMI 124
gi 2499383  34 . [16] .LVFVYH. [1] .RDIESPRAKLIIVATLMVPLVSISSYAGLASGLTVGFLQMP. [11] .SPWGRYLTWTFSTPMI 119
gi 1168614  4 . [16] .AVLAYG YTLVPEETRRKRYLLIAIPGIAIVAYALMALGFGSIQSEGH. [1] .VYVVRVVDULLTTPLN 83
gi 2499387  14 . [16] .LVFIAR. [1] .WVSVDQRQKFYIATINIAIAFVNYLSMALGFGVTITELG. [5] .IYYARYTDWLFITPLL 98
gi 2499366  7 . [16] .LVFIAR. [1] .WGETDPRQKFYIATILITIAIAFVNYLSMALGFGLTIVEFA. [5] .IYYARYSDWLFITPLL 91

```

NCBI

Conserved Domains

pfam01036.12 Bac_rhodopsin, with user query added

[+] Links: Bacteriorhodopsin.

[+] Statistics:

[+] Structure:

Show Structure

Program: Cn3D

Drawing: All Atoms

Aligned Rows: up to 5

[Download Cn3D]

Other Related Conserved Domains: C005524

Reformat Sequence Alignment Format: Compact Hypertext Row Display: up to 5 Color Bits: 2.0 bits Type Selection: the most similar member

```

1E12_A      8 . [16] .LVFVYHRTIRPGRPRLIWGATLMIPLVSISSYLGLLSGLTVGHIEMPAGH. [8] .SQWGRYLTWALSTPHILL 100
query      21 . [16] .FYFIVKGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVGSEM. [2] .IYYARYADWLFITPLLLL 107
1UAZ_A     15 . [16] .FYFIVKGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVGSEM. [2] .IYYARYADWLFITPLLLL 101
1MOK_A     22 . [16] .LVFLVKGVSDPAKKFYAITTLVPAIAFTNLSMLLGYGLTMVFPFGEQ. [2] .IYYARYADWLFITPLLLL 108
gi 2499387  14 . [16] .LVFIARGVSDQRQKFYIATINIAIAFVNYLSMALGFGVTITELGGEE. [2] .IYYARYTDWLFITPLLLY 100

```

CDD Descriptive Items

Name: Bac_rhodopsin

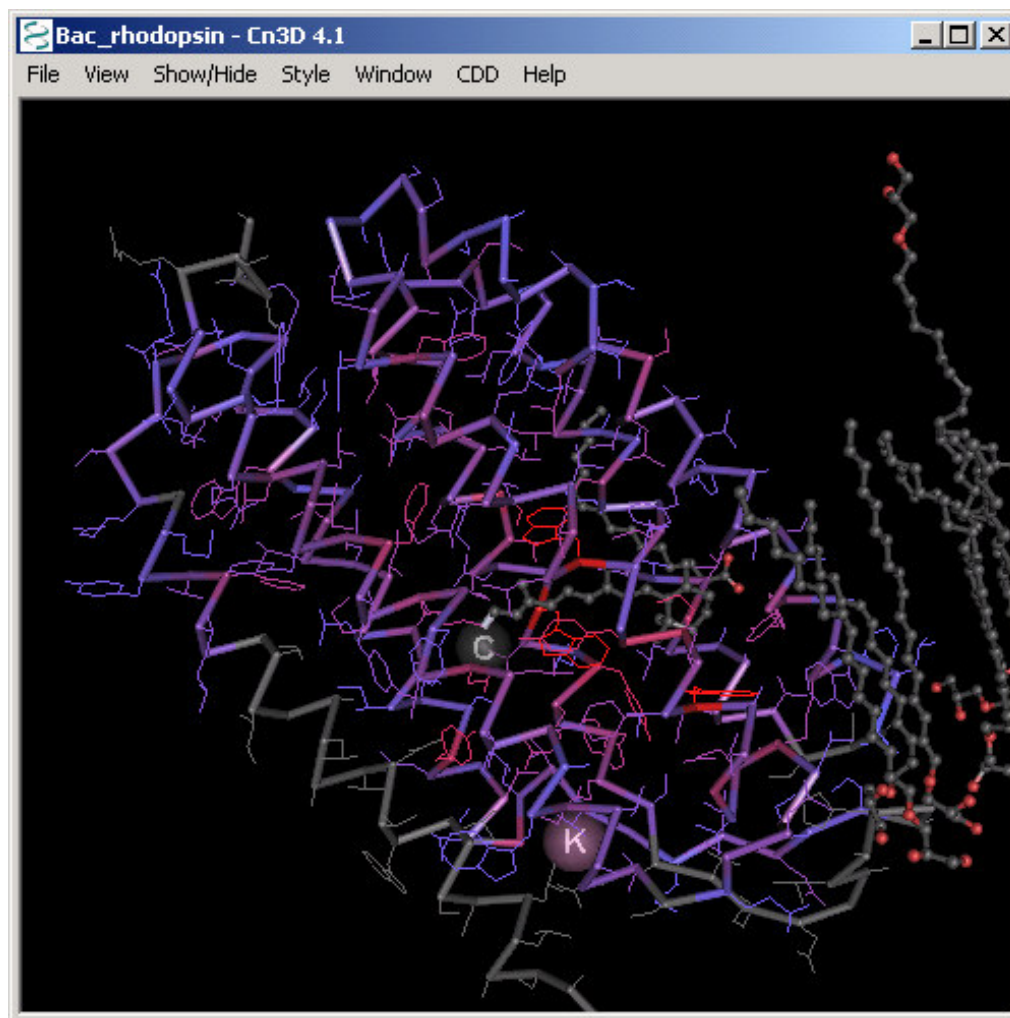
Bacteriorhodopsin.

Structure summary:

PDB 1E12 (MMDB 13348)

1E12_A: gi 8569313 ([Halobacterium salinarum] Chain A, Halorhodopsin, A Light-Driven Chloride Pump)

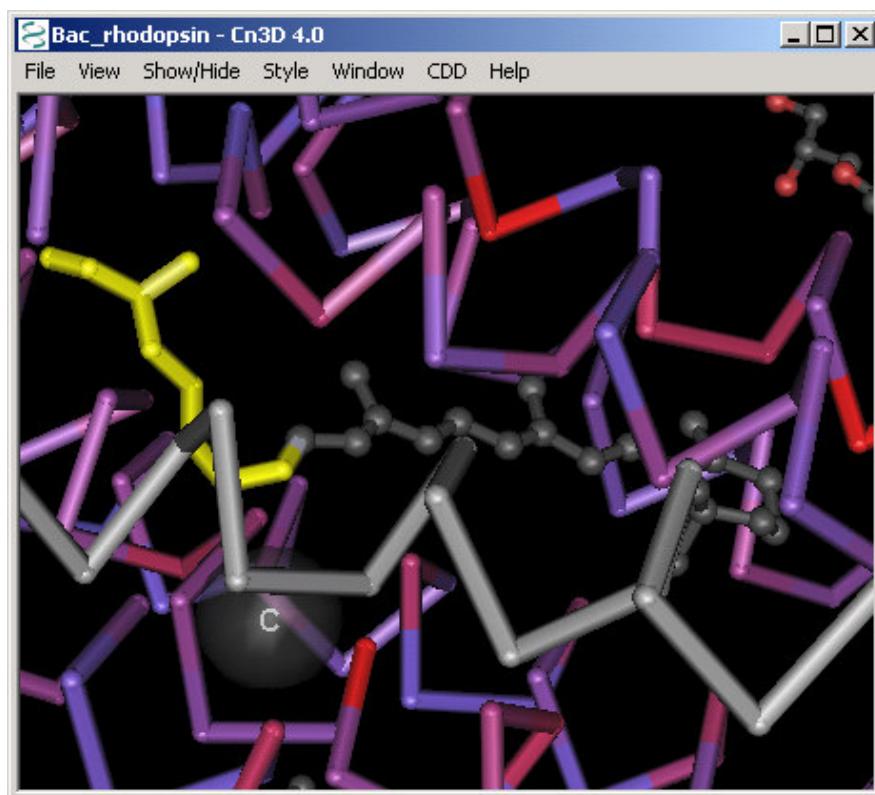
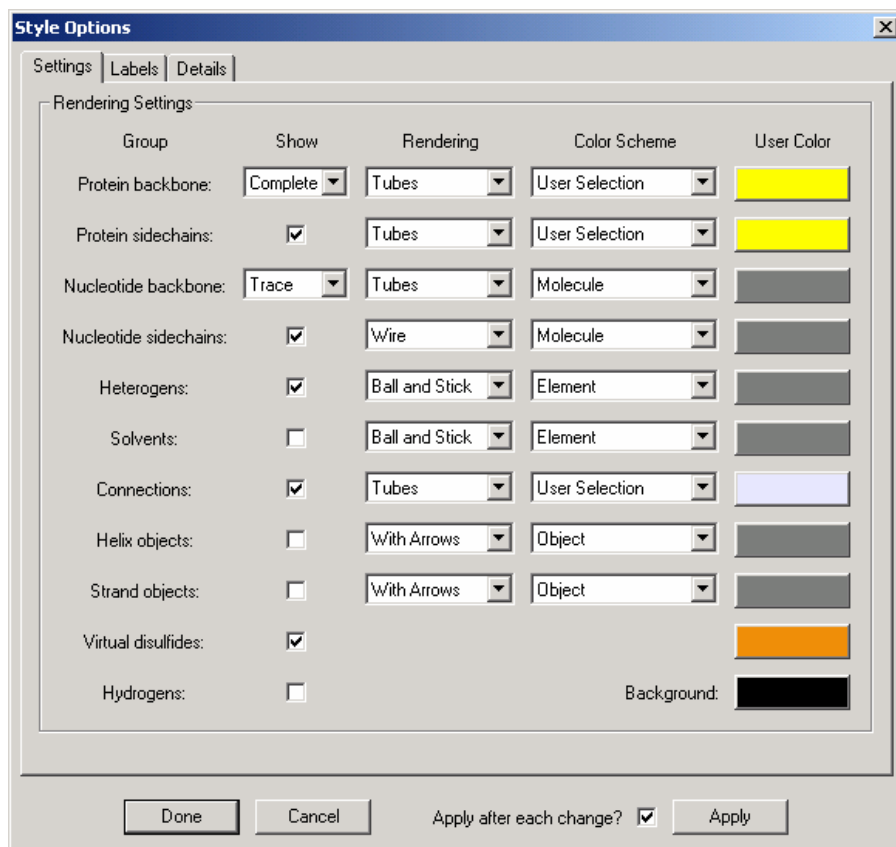
Show Annotations Panel Show References Panel Dismiss



Bac_rhodopsin - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

<i>LE12_A</i>	aglaailVFVYMGRTIRPGRPLIWGATLMIPLVSISSYLGLLSGLTVGMIEMP	aghal a ~ ~ ~ ~ ~ gem
<i>query</i>	mligtFYFIVKgWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG	sem ~ ~ ~ ~ ~
<i>IUAZ_A</i>	mligtFYFIVKgWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG	sem ~ ~ ~ ~ ~
<i>IMOK_A</i>	mligtLYFLVKgMGVSDPAKKFYAITTLVPAIAFTMYLSMLLGYGGLTMVPFG	geq ~ ~ ~ ~ ~
<i>gi 114809</i>	aglsiLLFVFMtRGLDDPRAKLI AVSTILVPVVSIASTYGLASGLTISVLEMP	aghfaegssvmlggee vdg
<i>gi 461609</i>	aglsiLLFVYMG RNVEDPRAQLI FVATLMVPLVSISSYTGLVSGLTVSFLEMP	aghal a ~ ~ ~ ~ ~ gqe
<i>gi 2499383</i>	agvviLLFVAMgRDIESPRAKLIWVATMLVPLVSISSYAGLASGLTVGFLQMP	pghal a ~ ~ ~ ~ ~ gqe
<i>gi 1168614</i>	ellatAVLAYG-YTLVPEETRKRYLLLLIAIPGIAIVAYALMALGFGSIOSEGH	a ~ ~ ~ ~ ~
<i>gi 2499387</i>		



NCBI Conserved Domains

pfam01036.12 Bac_rhodopsin, with user query added

Links: Statistics: Structure:

Show Structure

Program: Cn3D

Drawing: All Atoms

Aligned Rows: up to 5

[Download Cn3D]

Other Related Conserved Domains: C00524

Reformat Sequence Alignment

Format: Compact HyperText

Row Display: up to 5

Color Bits: 2.0 bits

Type Selection: the most similar members

1E12_A 8 . [16] .LVFVTRGTIRPGRPLINGATLMIFLVSISSTLGLLSGLTVGMENPAGH. [8] .SQNGRYLTWALSTPMILL 100

query 21 . [16] .FTFIVKGGWDTDEAREYYSITILVPGIASAAYLSFFGIGLTVQVQSEN. [2] .IYARYADWLFPTPLLL 107

1U42_A 15 . [16] .FTFIVKGGWDTDEAREYYSITILVPGIASAAYLSFFGIGLTVQVQSEN. [2] .IYARYADWLFPTPLLL 101

1BOK_A 22 . [16] .LVFLVKGWGGWDTDEAREYYSITILVPGIAATYLSFFGIGLTVQVQSEN. [2] .IYARYADWLFPTPLLL 108

gi_2499387 14 . [16] .LVFARGWGGWDTDEAREYYSITILVPGIAATYLSFFGIGLTVQVQSEN. [2] .IYARYADWLFPTPLLY 100

1E12_A 101 ALGLLADVDLSLFTVIAADICHCVTGLAAANT. [1] .SALLFRWAFYAFSCAFFVVLVSALVTDVAASASSA OT 171

query 108 DLALLARVDVSIOTLVGVDALHIVTGLVGLS HTPLARYTWLFTICHIVLVFLATSLRAAAKER. [2] .EV 179

1U42_A 102 DLALLARVDVSIOTLVGVDALHIVTGLVGLS HTPLARYTWLFTICHIVLVFLATSLRAAAKER. [2] .EV 173

1BOK_A 109 DLALLARVDVSIOTLVGVDALHIVTGLVGLS HTPLARYTWLFTICHIVLVFLATSLRAAAKER. [2] .EV 180

gi_2499387 101 DLALLAGADNTIYSLVGLVNIQTGLATLS. [6] .PAGAELVWGIISTOTFLVLLVFLSNLDRASEL. [2] .DL 178

NCBI Protein

Search Protein for 1E12A[ACCN]

Display GenPep Show 20 Send to

Range from begin to end Features: CDD Refresh

1: 1E12A Reports Chain A, Halorhodopsin. [gi:8569313]

Comment Features Sequence

LOCUS 1E12_A 253 aa linear BCT 06-APR-2000

DEFINITION Chain A, Halorhodopsin, A Light-Driven Chloride Pump.

ACCESSION 1E12_A

VERSION 1E12_A GI:8569313

DBSOURCE pdb: molecule 1E12, chain 65, release Apr 6, 2000;

deposition: Apr 6, 2000;

class: Ion Pump;

source: Mol_id: 1; Organism_scientific: Halobacterium Salinarum;

Strain: D2; Cellular_location: Membrane; Gene: Hop; Other details:

H. Sal. Strain D2 Was Constructed For Homologous Overexpression Of

Hr. See Also Heymann Et Al., Mol. Microbiol., Vo. 7, 623-630

(1993).;

Exp. method: X-Ray Diffraction.

KEYWORDS .

SOURCE Halobacterium salinarum

ORGANISM Halobacterium salinarum

Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

Halobacteriaceae; Halobacterium.

REFERENCE 1 (residues 1 to 253)

AUTHORS Havelka, W.A., Henderson, R. and Osterhelt, D.

TITLE Three-dimensional structure of halorhodopsin at 7 Å resolution

JOURNAL J. Mol. Biol. 247 (4), 726-738 (1995)

PUBMED 7723027

REFERENCE 2 (residues 1 to 253)

AUTHORS Osterhelt, D.

TITLE The structure and mechanism of the family of retinal proteins from

Halobacterium salinarum

Links

Related Structure

Related Sequences

3D Domains

Domain Relatives

PubMed

Structure

Taxonomy

NCBI Structure

Search Structure for 1E12

Display Summary Show 20 Send to

All: 1 Bacterial: 1 Eukaryotic: 0 Ligand: 1 NMR: 0 X-ray: 1

1: 1E12

Halorhodopsin, A Light-Driven Chloride Pump

[mmdbId:13348]

VAST, Links

[PubMed](#)
[BLAST](#)
[Structure](#)
[Taxonomy](#)
[OMIM](#)
[Help?](#)
[Cn3D](#)

VAST neighbors for: [MMDB 13348](#), [1E12 A](#)

Overview: There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced neighbor search controls. The second section is the VAST neighbor list itself.

View 3D Alignment
of
All Atoms
with
Cn3D
Display
Download Cn3D!

View Sequence Alignment
using
Hypertext
for
Selected
VAST neighbors

List
Medium redundancy
subset, sorted by
Aligned Length
in
Graphics
Graphics
Table
Download Asn1
Download Xml
Entrez

Advanced neighbor search

[PubMed](#)
[BLAST](#)
[Structure](#)
[Taxonomy](#)
[OMIM](#)
[Help?](#)
[Cn3D](#)

VAST neighbors for: [MMDB 13348](#), [1E12 A](#)

Overview: There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced neighbor search controls. The second section is the VAST neighbor list itself.

View 3D Alignment
of
All Atoms
with
Cn3D
Display
Download Cn3D!

View Sequence Alignment
using
Hypertext
for
Selected
VAST neighbors

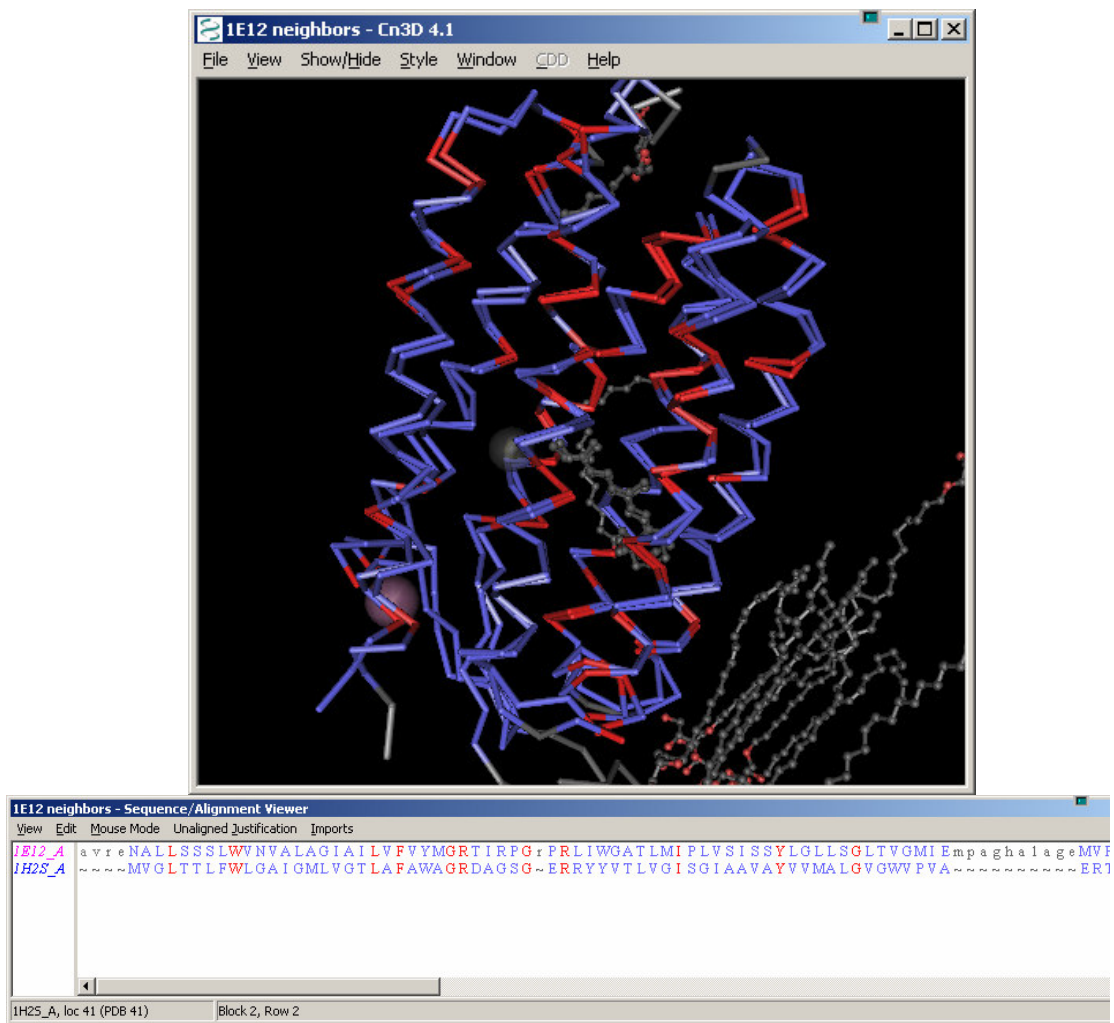
List
Medium redundancy
subset, sorted by
Aligned Length
in
Table

Advanced neighbor search

Total neighbors: 125; 18 representatives from the Medium redundancy subset displayed.

Click to: [Check All](#) [Uncheck All](#)

	PDB	C	D	Ali. Len	Score	E_Val	Rmsd	%Id	MMDB	Date	LHM	GSP	Description
<input type="checkbox"/>	1H2S	A		219	15.6	10e-15.9	1.4	27.4	11/2002	3.3	0.7		Molecular Basis Of Transmembrane Signalling By Sensory Rhodopsin li-Transducer Complex
<input type="checkbox"/>	1C3W	A		217	15.2	10e-15.1	1.6	33.6	03/2001	2.2	0.8		BacteriorhodopsinLIPID COMPLEX AT 1.55 A RESOLUTION
<input type="checkbox"/>	2F93	A		216	16.0	10e-17.0	1.3	27.8	05/2006	3.3	0.7		K Intermediate Structure Of Sensory Rhodopsin liTRANSDUCER Complex In Combination With The Ground State Structure
<input type="checkbox"/>	1XJO	A		209	11.9	10e-11.1	1.7	26.3	11/2004	4.5	0.8		Anabaena Sensory Rhodopsin



Problem 2

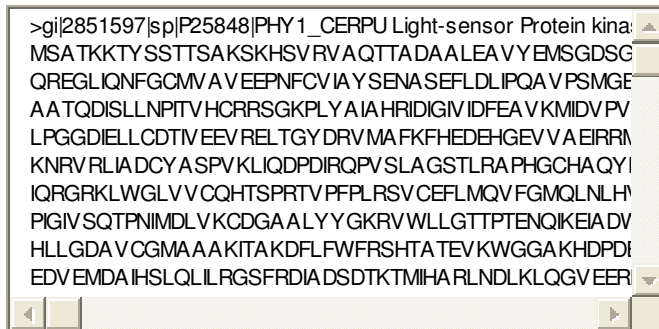
In this problem, we will follow these steps:

- Identify conserved domain(s) present in a protein.
- Search for other proteins containing similar domain(s).
- Explore a 3D modeling template for the query sequence.
- Find distant sequence homologs that may not be identified by BLAST.

NCBI's Conserved Domain Search allows you to match your protein sequence to a library of conserved protein domains, generate a multiple sequence alignment based on this match, and explore 3D modeling templates for your sequence. Click on the CDD link provided below,

CDD

paste the following protein sequence in the CD-Search query box and run the search.

A screenshot of a web browser showing a protein sequence in a text box. The sequence is: >gi|2851597|sp|P25848|PHY1_CERPU Light-sensor Protein kina: MSATKKTY SSTTSAKSKHSVRVAQTTADAALEAVYEMSGDSG QREGLIQNFGCMVAVEEPNFCVIA YSENA SEFLDLIPQAVPSMGE AATQDISLLNPITVHCRRSGKPLYAIAHRIDIGIVIDFEAVKMIDVPV LPGGDIELLCDTV EEV RELTG YDRVMAFKFHEDEHGEVVAEIRRN KNRVRLIADCYASPVKLIQDPDIRQPVSLAGSTLRAPHGCHAQYI IQRGRKLWGLVV CQHTSPRTVPFPLRSVCEFLMQVFGMQLNLH) PIGIVSQTPTNIMDLVKCDGAALYYGKRVWLLGTTPTENQIKEIADV HLLGDAVCGMAAAKITAKDFLWFRSHTATEVKWGGAKHDPDI EDVEMDAIHSLQLILRGSFRDIA DSDTKTMIHARLNDLKLQGV EER. The text box has a scrollbar on the right and navigation buttons at the bottom.

- What are the domains present in this protein?
(Select the "Full Result" radio button to display all of the domains.)

-Suppose, we are interested in the serine/threonine protein kinase domain.
Obtain more information about it by searching in [NCBI's Bookshelf](#)

- Go back to the CD-Search results page. Obtain a list of proteins with similar domain architecture by clicking on the "Search for similar domains architectures" button. To display the records, click on the links to the subsets of sequences and from there on the "Look up Sequences in Entrez". Change the display from "Summary" to "FASTA".

- Go back to the CD-Search results page. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the graphic representation of the serine/threonine kinase domain from CDD (CDD|00180). Use the "Row Display" list box pull down menu to specify "up to 5" sequences and reformat sequence alignment. Invoke Cn3D

with a display of a 3D modeling template and a multiple sequence alignment including your query sequence by pressing the "Show Structure" button.

To show only one top structure, click on the down arrow key. For better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a serine/threonine protein kinases active-site signature (IIHRDLKSMNILV) where K is the ATP binding site. Identify these residues in the query protein and highlight the corresponding lysine residue in the first protein sequence.

Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click the Protein Side chains on. Click on the "Done" button. To zoom in, press z on the keyboard. Note the heterogen near the lysine residue.

D. To obtain the structure neighbors for the serine/threonine protein kinase protein, first click on the structure entry, 1JNK, on the CD-Browser page. Then click Links → Structure on the top right, then on 1JNK again in the Entrez Structure page, and finally on the chain graphic. To view neighbors with 1JNK, select one or more of the check boxes next to the structure neighbors and view by clicking on the "View 3D Structure" button.